

# How to Calculate a Confidence Interval for a Regression Intercept: A Step-by-Step Guide

Authored by  
**stats writer**

December 2, 2025

## RECOMMENDED CITATION

stats writer (2025). *How to Calculate a Confidence Interval for a Regression Intercept: A Step-by-Step Guide*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=103589>

Calculating a Confidence Interval for a Regression Intercept is a crucial step in understanding the uncertainty associated with statistical modeling results. This process primarily relies on the use of the Standard Error of the intercept estimate, typically derived from a t-test framework. The resulting interval defines a range of plausible values for the true population intercept (often denoted as  $\beta_0$ ), usually set at a 90%, 95%, or 99% confidence level, providing a robust measure of reliability for this key parameter.

## Understanding Simple Linear Regression and its Components

Simple linear regression (SLR) serves as a foundational statistical technique used to quantify the linear relationship existing between one predictor variable (independent variable,  $x$ ) and a response variable (dependent variable,  $y$ ). The fundamental goal of SLR is to determine the line that minimizes the sum of squared residuals, thus providing the "best fit" for the dataset.

This best-fit line is mathematically represented by the equation that defines the predicted response ( $\hat{y}$ ):

$$\hat{y} = b_0 + b_1x$$

In this model, each coefficient carries significant interpretative weight:

$\hat{y}$ : The estimated or predicted value of the response variable.

$b_0$ : The Regression Intercept, representing the estimated average value of the response variable ( $y$ ) when the predictor variable ( $x$ ) is exactly zero.

$b_1$ : The slope of the regression line, indicating the average change in the response variable associated with a one-unit increase in the predictor variable.

$x$ : The observed value of the predictor variable.

While researchers often focus intensely on the slope ( $b_1$ ) to understand the magnitude and direction of the relationship, the intercept ( $b_0$ ) is essential for completing the model. Understanding the uncertainty around the intercept requires constructing a Confidence Interval for the true population intercept,  $\beta_0$ .

## The Formula for the Intercept Confidence Interval

To quantify the reliability of the estimated intercept ( $b_0$ ), we must calculate its corresponding confidence interval. This interval uses the point estimate ( $b_0$ ), the appropriate critical value from the t-distribution, and the Standard Error of the intercept estimate ( $se(b_0)$ ).

The general formula for calculating the Confidence Interval for the true population intercept ( $\beta_0$ ) is:

### Confidence Interval for $\beta_0$ : $b_0 \pm t_{\alpha/2, n-2} * se(b_0)$

Here,  $t_{\alpha/2, n-2}$  is the critical t-value determined by the desired confidence level ( $\alpha$ ) and the degrees of freedom ( $n-2$ , where  $n$  is the sample size). The  $se(b_0)$  term, the standard error of the coefficient, measures the expected sampling variability of the intercept estimate across different samples. A larger standard error results in a wider, and thus less precise, confidence interval.

### Practical Example: Setting up the Regression Model (Data and R Code)

To illustrate the calculation, let us analyze a dataset where we fit a simple linear regression model. Suppose we collect data on 15 students, using the hours studied as the predictor variable and the resulting exam score as the response variable. This scenario allows the intercept to be potentially interpretable, as zero hours studied is a plausible value.

The data collected for these 15 students is structured as follows:

Hours Studied	Exam Score
1	64
2	66
4	76
5	73
5	74
6	81
6	83
7	82
8	80
10	88
11	84
11	82
12	91
12	93
14	89

We use the statistical software R to fit the simple linear regression model. The code below defines the dataset and executes the linear model fitting process using the `lm()` function:

```
#create data frame
```

```
df <- data.frame(hours=c(1, 2, 4, 5, 5, 6, 6, 7, 8, 10, 11, 11, 12, 12, 14),
```

```
score=c(64, 66, 76, 73, 74, 81, 83, 82, 80, 88, 84, 82, 91, 93, 89))
```

```
#fit simple linear regression model
```

```
fit <- lm(score ~ hours, data=df)
```

```
#view summary of model
```

```
summary(fit)
```

Call:

```
lm(formula = score ~ hours, data = df)
```

Residuals:

```
Min 1Q Median 3Q Max
```

```
-5.140 -3.219 -1.193 2.816 5.772
```

Coefficients:

```
Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept) 65.334 2.106 31.023 1.41e-13 ***
```

```
hours 1.982 0.248 7.995 2.25e-06 ***
```

```
---
```

```
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 3.641 on 13 degrees of freedom

Multiple R-squared: 0.831, Adjusted R-squared: 0.818

F-statistic: 63.91 on 1 and 13 DF, p-value: 2.253e-06

## Analyzing the Regression Output

The output generated by the R summary provides all the necessary components for calculating the confidence interval for the intercept. The core information extracted from the "Coefficients" table is critical for our calculations:

**Intercept Estimate (\$b\_0\$):** 65.334

**Standard Error of the Intercept (\$se(b\_0)\$):** 2.106

**Degrees of Freedom (\$n-2\$):** 13 (derived from \$n=15\$ students minus 2 parameters estimated)

Based on these estimates, the fitted simple linear regression model relating hours studied to exam score is:

Score = 65.334 + 1.982 \* (Hours Studied)

The intercept value of 65.334 indicates that, according to the model, a student who studies for zero

hours is predicted to achieve an average score of 65.334. We now proceed to determine the range around this estimate that is likely to contain the true population intercept.

## Step-by-Step Calculation of the Confidence Interval

We aim to calculate the 95% Confidence Interval for  $\beta_0$ . This requires finding the appropriate t critical value ( $t_{\alpha/2, n-2}$ ) for a two-tailed test with  $\alpha = 0.05$  (or  $\alpha/2 = 0.025$ ) and 13 degrees of freedom.

Consulting a t-distribution table (or using statistical software) for  $df=13$  and  $p=0.025$  (one-tail area), we find the critical t-value is  $t_{\{0.025, 13\}}$  approx 2.1604.

Now we apply the confidence interval formula:

95% C.I. for  $\beta_0$ :  $b_0 \pm t_{\alpha/2, n-2} * se(b_0)$

95% C.I. for  $\beta_0$ :  $65.334 \pm t_{0.05/2, 15-2} * 2.106$

95% C.I. for  $\beta_0$ :  $65.334 \pm 2.1604 * 2.106$

95% C.I. for  $\beta_0$ :  $65.334 \pm 4.549$

95% C.I. for  $\beta_0$ :

We conclude that we are 95% confident that the true population mean exam score for students who study for zero hours falls between 60.78 and 69.88. This interval provides a statistical measure of certainty around the model's baseline prediction.

## Critical Caveats: When Intercept Interpretation Fails

It is vital to recognize that calculating a Confidence Interval for the intercept is only meaningful if the intercept itself has a relevant interpretation within the context of the study. The intercept,  $\beta_0$ , only represents the predicted value of  $y$  when  $x=0$ . If  $x=0$  is a non-existent, impossible, or illogical value for the predictor variable, then interpreting  $\beta_0$  is meaningless, and consequently, calculating its confidence interval is statistically unwarranted.

Consider a regression model predicting the average points per game of a basketball player based on their height. A player cannot have a height of zero feet. Therefore, interpreting the Regression Intercept in this context--the predicted score of a player zero feet tall--is a case of extrapolation far outside the scope of the observed data, rendering the calculation of its confidence interval superfluous.

Numerous predictor variables commonly used in regression analysis cannot logically take on a value of zero. Attempting to calculate or interpret the intercept CI in these scenarios leads to statistical nonsense. Examples of such variables include:

Square footage of a house.

Length of a car.

Weight of a person.

In all these instances, if these variables are used as predictors, the  $x=0$  condition is not met in reality, and researchers should focus their interval estimation efforts on the slope coefficient ( $b_1$ ) instead, or center the predictor variable to make the intercept interpretable (i.e., transform  $x$  so that  $x=0$  represents the mean value of the predictor).

## Conclusion and Further Resources

While the methodology for calculating a confidence interval for the regression intercept is straightforward--relying on the estimate, the Standard Error, and the t critical value--it is essential to first assess the contextual relevance of the intercept itself. Only when the predictor variable can plausibly take on a value of zero should this interval be calculated and interpreted. This due diligence ensures that statistical rigor is maintained and that the resulting interpretations are grounded in reality.

For those interested in delving deeper into related concepts, the following tutorials provide additional information about linear regression modeling and inference: